Purification and Labeling of human MLH1-PMS1

Undergraduate Research Thesis

Presented in Partial Fulfillment of the Requirements for Graduation "with Honors Research Distinction in Microbiology" in the undergraduate colleges of The Ohio State University

By

Rachel Lew

The Ohio State University

May 2020

Project Advisor: Dr. Richard Fishel, Department of Cancer Biology and Genetics

Abstract

A common mechanism that many species have to repair errors in DNA replication is DNA mismatch repair. Mutations in the proteins involved in this mechanism correspond with an accumulation of mismatched nucleotides. The accumulation of mismatched nucleotides can lead to cancers or DNA instability (3,5). For example, Huntington's Disease, a trinucleotide repeat disease, results after DNA mismatch repair proteins cause trinucleotide repeat-induced DNA instability (5). Furthermore, Lynch syndrome corresponds to a defective mismatch repair system and people who inherit this disease have a higher risk of developing colorectal cancer (3).

The two critical proteins in the mismatch repair system are MutS and MutL. MutS searches the DNA for a mismatch, and once a mismatch is recognized, it loads MutL. Once both of these proteins are on the DNA, the mismatch repair can proceed (3). Of the human MutL homologs, MLH1-PMS1 has not been studied comprehensively because it is a much weaker genetic mutator than the other homologs and it does not have the latent endonuclease activity that MLH1-PMS2 has (1). However, evidence suggests that MLH1-PMS1 functions in MSH2-MSH3 mismatch repair.

A protocol to purify and label this protein using chromatography and sortase labeling is presented so that the purified protein can be used in single molecule experiments. Further investigation of this protein will likely provide insight on the mismatch repair mechanism, Lynch syndrome, and trinucleotide repeat diseases such as Huntington's disease.

Acknowledgements

I would like to express my sincere gratitude to several individuals who have supported me throughout my undergraduate career. First, I would like to thank my advisor, Dr. Richard Fishel, for giving me the opportunity to pursue this research project and for supporting my goals and research interests. I would also like to thank James London for all the time he spent training me, encouraging me, pushing me to work hard, and giving me advice. Without their support and guidance, this project would not have been possible. Secondly, I would like to thank all members of the Fishel and Yoder labs for answering any questions that I had at any point and for teaching me to become a better scientist. I am also extremely grateful to Dr. Natividad Ruiz and Dr. Chad Rappleye for agreeing to be on my thesis defense committee and for helping me build my knowledge of microbiology outside of the lab. Finally, I would like to thank my friends and family who have provided me guidance, supported my endeavors, and have always motivated me to keep going.

Introduction

Mismatch repair (MMR) is a highly conserved process that identifies and fixes nucleotide mismatch errors that propagate during the process of DNA replication. These nucleotide mismatches are attributed to a variety of causes, including polymerase misincorporation errors, recombination between heteroallelic parental DNAs, and chemical/physical damage to nucleotides. Failure of MMR corresponds with an abundance of mispaired nucleotides, culminating in an accumulation of genetic mutations that can cause DNA instability or cancer (3,5,7). Thus, the importance of mismatch repair on the development of diseases cannot be understated.

Huntington's Disease, which causes the breakdown of nerve cells in the brain, corresponds to DNA instability due to the accumulation of trinucleotide repeats (5). In particular, it has been demonstrated that MSH2 is accountable for almost 100% of (CAG)·(CTG) repeat expansions in this disease and mutations in MSH2 eliminate the presence of these expansions (11). Furthermore, studies have shown that mutations in MSH3, rather than MSH6, eliminates the presence of trinucleotide repeat expansions as well (13). It is estimated that the possibility of passing Huntington's Disease onto progeny is 50% and that overall, 5.70 out of 10,000 people are affected (5,10). On the other hand, Lynch syndrome, the most common hereditary colorectal cancer condition which is estimated to affect about 1 in 370

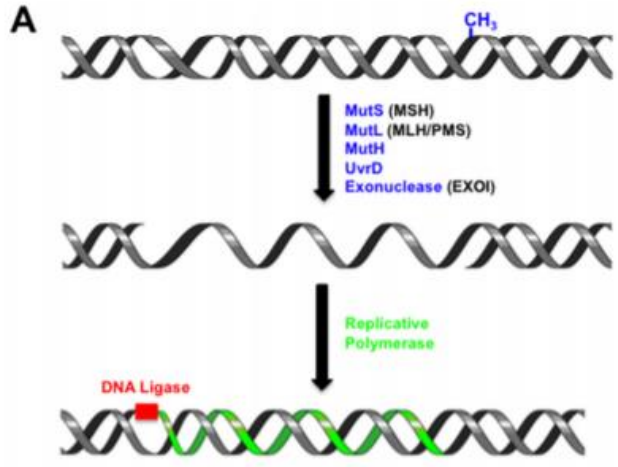

Figure 1. Basic Mismatch Repair Scheme from Fishel, R. (2015). Mismatch Repair. The Journal of Biological Chemistry, 290(44), 26395–26403

to 1 in 2000 people (9), is caused by autosomal dominant mutations in the mismatch repair genes and results in the accumulation of nucleotide alterations (3,7,12). As such, there is considerable value, when investigating Huntington's Disease, Lynch Syndrome, and a host of other diseases, in attempting to understand the nature of mismatch repair and how faulty functioning among its many protein components has substantial consequences for human health.

In *Escherichia coli*, MutS is an ADP-bound homodimer that acts as a short-lived sliding clamp to intermittently search DNA. Once a mismatch is recognized, MutS binds to the mismatch and releases ADP. It then binds ATP and forms a stable-long lived sliding clamp that randomly slides along the DNA until the ATP is hydrolyzed by the weak ATPase activity of MutS. This releases the clamp from the DNA and returns it to the ADP-bound state so that it can continue its search of the DNA (3,7). MutL, also a homodimer, is recruited to the DNA by ATP-bound sliding MutS clamps where it binds ATP to form a second independent sliding clamp that can disassociate and re-associate with MutS. As a sliding clamp, MutL can then recruit MutH to nick the strand of DNA at hemi-methylated d(GATC) sites. This nick allows the proteins associated with the excision reaction to target the newly synthesized daughter strand and repair the mismatch. The MutL sliding clamp is released from the DNA upon ATP hydrolysis by the weak ATPase function of MutL(3,7). Other studies have shown that while this mechanism is similar in eukaryotes, such as yeast and humans, there are some differences and it is less understood (4). However, the conservation of processes such as recognition, removal, and re-synthesis, is well-documented (3,6).

E. coli Protein	Summary
MutS	ADP-bound homodimer that searches the DNA for a mismatch. Once a mismatch is recognized, MutS binds to the mismatch and releases ADP. It then binds ATP and can recruit MutL to the DNA (3, 7).
MutL	Once recruited to the DNA, MutL binds ATP forms an independent sliding clamp that interacts with MutS. MutL can recruit MutH (3, 7).
MutH	Nicks the DNA at hemi-methylated d(GATC) sites (3,7).

Table 1. Summary of mismatch repair protein functions in *E. coli*

In humans, there are two MutS homologs that are capable of recognizing mismatches: MSH2-MSH6 and MSH2-MSH3. MSH2-MSH6 recognizes base-base mismatches or small insertion/deletion loops of 1-2 nucleotides, while MSH2-MSH3 recognizes larger insertion deletion loops of 1-14 nucleotides (2). Once the mismatch is recognized, the MutS homolog recruits and forms a complex with the MutL homolog, MLH1-PMS2 (3). In contrast to the prokaryotic *E. coli* mismatch repair system, the eukaryotic system does not have a homologous protein to MutH (6). However, the endonuclease activity of MLH1-PMS2, known to be activated by PCNA, is thought to do the nicking (3). Furthermore, after the DNA incision, the nicked DNA strand is removed. It is hypothesized that EXO1 is recruited to promote excision of the mismatched DNA. However, other mechanisms are possible. The mismatch repair pathway is completed by DNA polymerase and DNA ligase (6).

Current literature has not elucidated a specific role for MLH1-PMS1 in MMR. In the past, PMS1 has not been studied as comprehensively as PMS2 because PMS1 lacks the endonuclease activity that PMS2 has. Furthermore, mice lacking PMS1 are not cancer prone, suggesting an alternate role for PMS1 in the mismatch repair system (1). Through co-immunoprecipitation, it has been shown that MLH1 and PMS1 have a very high affinity for each

other but molecules that interact with PMS1 are mainly proteins belonging to the ubiquitylation pathway, leading to a hypothesis that PMS1 could be post-translationally modified by ubiquitin. This polyubiquitylation may target PMS1 for proteasome-mediated degradation which might make MLH1 more available for PMS2 for mismatch repair (1). However, because many proteins get ubiquitylated in cells, the fact that PMS1 is ubiquitylated for the sole purpose of freeing up MLH1 is unlikely. Another hypothesis that seems to be more valid is that PMS1 participates in MSH2-MSH3 mismatch repair, which functions in repair of larger insertion-deletion loops. This claim is based on one of the MutL homologs in yeast, MLH1-MLH3, that is hypothesized to participate in MSH2-MSH3 mismatch repair because mutations in this protein cause an increase in the number of frameshift mutations (6). Thus, the importance of human MLH1-PMS1 seems to lie in reparation of mismatches in trinucleotide repeat diseases such as Huntington's Disease or Myonic Dystrophy.

Human Protein	Summary
MSH2-MSH6	Recognizes base-base mismatches and small insertion/deletion loops of 1-2 nucleotides (2)
MSH2-MSH3	Recognizes large insertion/deletion loops of 1-14 nucleotides (2)
MLH1-PMS2	Has a latent endonuclease activity, that is activated by PCNA, which is thought to do the nicking (3)
MLH1-PMS1	Function currently unknown but is hypothesized to participate in MSH2-MSH3 mismatch repair (1)

Table 2. Summary of mismatch repair proteins in humans.

Overview

In this paper, a method of labelling and purifying MLH1-PMS1 will be presented so that this procedure can be used for future single-molecule experiments. The protein was labeled with a Sulfo-Cy5 dye using a sortase-mediated reaction and was purified using an FPLC with a Ni-NTA, Q-Sepharose, and PBE column.

Materials and Methods

Subcloning and Expression

A his6-tag and the sortase recognition site, CLPETGG, were added to the N-terminal end of PMS1 though PCR. The gene was then cloned into pFastBac through restriction digestion with BamHI and KpnI followed by ligation. The plasmid containing PMS1 was then transformed into *E. coli* XL10 Gold cells. Sequencing results confirmed the presence of the gene in the plasmid. In preparation of expression in Sf9 insect cells using guidelines from the Bac-to-Bac Baculovirus Expression System manual (Life Technologies), the gene was then transformed into DH10Bac *E. coli*. Blue/white selection was used to identify colonies that contained the recombinant bacmid. The Sf9 insect cells were transfected with the purified recombinant bacmid to produce a baculoviral stock, which was titered up three times. Two viral stocks were used to express the recombinant protein in Sf9 insect cells: his-srt-PMS1 and wildtype MLH1. For a typical expression, four 200 mL cultures were expressed at 27°C with 1 mL of each viral stock. The cell pellets were collected 48 hours later and stored at -80°C in storage buffer (25 mM HEPES pH 7.8, 100 mM NaCl, 10% Glycerol, 10 mM Imidazole pH 8.0) for later use.

Purification and Labelling

The following buffers were used in the full purification of MLH1-PMS1.

Ni Buffer A: 25 mM HEPES pH 7.8, 100 mM NaCl, 10% Glycerol, 10 mM Imidazole pH 8.0

Ni Superloop Buffer : 25 mM HEPES pH 7.8, 800 mM NaCl, 10% Glycerol, 10 mM Imidazole pH 8.0

Ni Buffer B : 25 mM HEPES pH 7.8, 100 mM NaCl, 10% Glycerol, 100 mM Imidazole pH 8.0

Q-Sepharose Buffer A: 25 mM HEPES pH 7.8, 10% Glycerol

Q-Sepharose Buffer B: 25 mM HEPES pH 7.8, 1 M NaCl, 10% Glycerol

Dialysis Storage Buffer: 25 mM HEPES pH 7.8, 150 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 20% Glycerol

Four cell pellets were thawed on ice at 4°C and centrifuged at 41000 rpm for 1 hour at 4°C. The supernatant was then injected onto a pre-equilibrated 3 mL Ni-NTA column at 0.15 mL/min using an FPLC and washed with 40 mL of *Ni Superloop Buffer*. The protein was then washed with an additional 10 mL of *Ni Buffer A* to remove the high salt and stepped to *Ni Buffer B* for 10 mL to elute 10 mL of the protein in 0.2 mL fractions. A 10% SDS-PAGE gel was run and it was determined that the protein was only semi-pure.

The semi-pure protein was labelled in a reaction containing 5x sortase, 115x Sulfo-Cy5 dye, 13 nmol of protein, and 10 mM $CaCl_2$. The labeling reaction was incubated at 4°C for 30 minutes. 20 mM EDTA was used to stop the reaction. The labelled protein was then injected onto a pre-equilibrated 2 mL Q-Sepharose column at 0.15 mL/min and washed for 50 mL with 10% *Q-Sepharose Buffer B* to prevent non-specific binding. A gradient to 60% *Q-Sepharose Buffer B* over 10 mL was used to elute the protein in 0.2 mL fractions. A 10% SDS-PAGE gel was run to determine the purity. However, while the protein bands for MLH1 and PMS1 appeared on the gel, the protein was still only semi-pure.

Next, the protein was injected onto a pre-equilibrated 2 mL PBE column at 0.15 mL/min to remove any further impurities and to obtain the best concentration. The protein was washed for 40 mL at 10% *Q-Sepharose Buffer B* to prevent non-specific binding and was eluted at 60% *Q-Sepharose Buffer B* for 10 mL in 0.2 mL fractions. Another 10% SDS-PAGE gel was run, and it was determined that the protein was pure with the exception of a 30 kDa band of unknown

identity. The protein was dialyzed with *Dialysis Storage Buffer* and frozen for use in future single-molecule experiments.

Results

An image of the 10% SDS-PAGE gel is not available for the Ni-NTA purification or the Q-Sepharose purification. However, after both purifications, it was clear that the purity was still questionable. After purification using the PBE column, the major protein peak was over fractions 5, 6, and 7 (see *Figure 2)*. The proteins were visualized on a 10% SDS-PAGE gel and there still seems to be some impurities (see *Figure 3*), but PMS1 (105.8 kDa) and MLH1 (84.6 kDa) are clearly present. Based on *Figure 3,* fraction 5 was chosen for future single molecule experiments because it seemed to have the highest amount of protein and the least amount of the unknown 30 kDa band. The specificity of labeling is shown in *Figure 4.* There seems to be some non-specific labeling. However, the majority of the protein labeled is PMS1. Thus, all three column purifications resulted in 4.8 uM of pure protein that was 67% labeled (see *Table 3*). Further analyses will need to be done on the band of about 30 kDa, whose identity is currently unknown, to determine what protein it is.

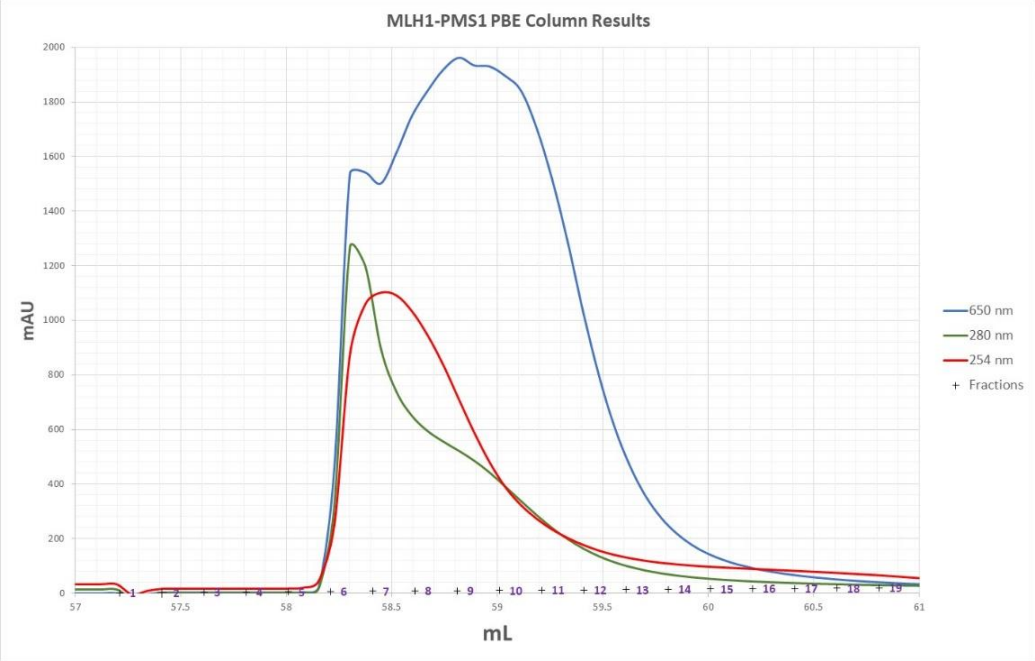

Figure 2. Chromatograph of PBE column purification. The major protein peak is over fractions 5, 6, and 7.

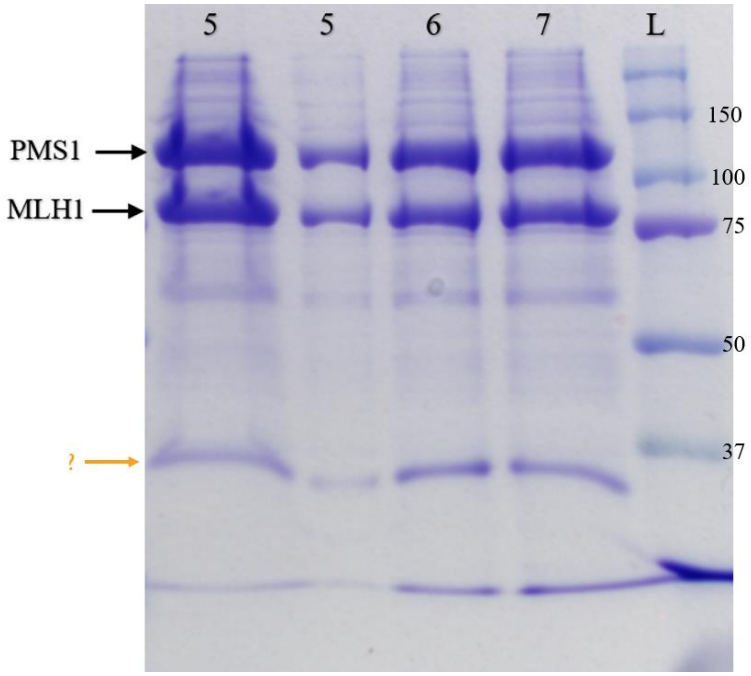

Figure 3. Coomassie stain of MLH1-PMS1 after PBE column. From left to right: Fraction 5, Fraction 5, Fraction 6, Fraction 7, Ladder. An unknown 30 kDa band is also present (orange).

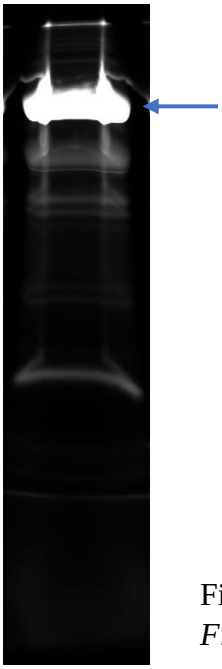

Figure 4. Cy5 Scan of 10% SDS-PAGE gel in *Figure 3*. Shown here is Lane 1, Fraction 5.

Protein	A280	Concentration of Cy5 Dye	Total Pure Protein	Percent Labeled
MLH1-PMS1	0.683	3.22 uM	4.8 uM	67%

Table 3. Total Pure Protein and Percent Labeled Protein

Discussion

Three columns (Ni-NTA, Q-Sepharose, and PBE) were used in order to purify MLH1-PMS1, a protein hypothesized to be involved in MSH2-MSH3 mismatch repair, for single molecule experiments. Criteria for column selection was primarily based on previous protocols in the lab for purifications with the other MutL homolog, MLH1-PMS2. The Ni-NTA column was used for the initial purification of the protein as the his-tag binds to the nickel with high affinity. The protein was loaded in low salt and 10 mM imidazole to prevent non-specific

binding to the column. The intention of the subsequent high-salt wash was to further disrupt non-specific binding and remove some impurities to obtain a purer protein product in the end. Elution of the protein off the column required an increase in the concentration of imidazole, which competes for binding to the nickel, to improve efficiency and more importantly obtain the highest concentration of protein in the smallest number of fractions possible.

The labelling was carried out using a sortase-mediated reaction with a Sulfo-Cy5 dye, a method that has been used in the lab with favorable results because it specific, produces high labeling yields, and can be carried out under physiological conditions (8). The Sulfo-Cy5 peptide was prepared using a HPLC. In this reaction, sortase recognizes the CLPETGG motif at the N-terminus end of PMS1 and cleaves the peptide bond between the threonine and the glycine residues using a catalytic cysteine. This generates a thioacyl intermediate. Subsequent nucleophilic attack by the N-terminus of PMS1 for the C terminus of the Sulfo-Cy5 peptide resolves the intermediate, which results in the formation of a covalent bond between the peptide probe and the protein (8). The conditions used for the reaction were determined in an effort to achieve the highest labelling possible and to push the equilibrium of the reaction towards product formation. A labelling efficiency of 67% was obtained, with back reactions from the final product and the reversible nature of the sortase reaction contributing to the decreased labelling efficiency (8).

For better separation of the protein from other impurities, a Q-Sepharose anion exchange column was used. The protein was once again loaded in low salt to prevent non-specific binding and eluted at high salt. No other reagents were added to the buffers except for 25 mM Hepes pH 7.8 to maintain physiological conditions and 10% glycerol, essential for maintaining protein stability. A gradient was employed to elute the protein so that the protein of interest would come

off the column at a different concentration from the non-essential proteins. Following this step, the protein was loaded onto a PBE column to enhance its concentration. The same Q-Sepharose buffer stocks were used to purify this protein because the PBE column works in a similar fashion, meaning that the protein elutes off the column at high salt concentrations because it is an anion exchange column. In this case, the protein was stepped off the column so the highest concentration of protein could be obtained. After completion of this purification, the protein was ~90% pure but the presence of impurities was still noted, highlighting a need to revisit the protocol in future experiments to improve yield and purity.

The protein was relatively pure with the exception of a 30 kDa band appearing at the bottom of the gel. One hypothesis is that the band is PCNA (30 kDa), which ensures the processivity of DNA polymerase, because of PMS1's similarity to PMS2 (3). Furthermore, a study involving co-immunoprecipitation experiments has shown that besides MLH1, PMS1 also interacts with RFC, a protein known to load PCNA onto the DNA during re-synthesis (1). The true identity of the band, however, is currently unknown and further analyses involving mass spectrometry, as well as confirmation with existing standard protein weights, will need to be done to determine which protein it is.

While an important step to improving the purification protocol is determining the identity of the 30 kDa band, there are other ways in which better separation and concentration of the protein can be achieved. For example, using a shallower gradient with the PBE column could yield a higher concentration of the PMS1 while washes with slightly higher salt concentrations could decrease the amount of non-specific binding obtained. However, caution must be taken when increasing the salt concentration so as not to prematurely elute the protein. Furthermore, higher labelling efficiencies would yield better results in single-molecule experiments. Studies

have shown that sortase labeling can achieve labeling efficiencies as high as 90% (8). So, increasing the amount of protein used or putting the label on the C-terminus could serve as viable methods to improve the amount of labelled protein.

Conclusion

MLH1-PMS1 was labelled with a Sulfo-Cy5 dye on the N-terminus and was purified using a Ni-NTA, Q-Sepharose, and PBE column. From this purification protocol, 4.8 uM of protein was obtained that was 67% labelled. An unknown 30 kDa band, hypothesized to be PCNA, was also purified alongside MLH1-PMS1. However, the true identity of the band will need to be determined in the future. The purified protein will be used for single-molecule experiments to study the interaction between MSH2-MSH3 and MLH1-PMS1 as well as any other proteins that this protein interacts with such as PCNA. These experiments can be used to elucidate the mechanism for trinucleotide repeat diseases such as Huntington's Disease or even Lynch syndrome. With more insight into how this disease progresses, the odds of developing and discovering novel prevention methods and treatments are higher.

References

1. Cannavo, E., Gerrits, B., Marra, G., Schlapbach, R., & Jiricny, J. (2006). Characterization of the Interactome of the Human MutL Homologues MLH1, PMS1, and PMS2. *Journal of Biological Chemistry*, *282*(5), 2976–2986. doi: 10.1074/jbc.m609989200
2. Dowen, J. M., Putnam, C. D., & Kolodner, R. D. (2010). Functional Studies and Homology Modeling of Msh2-Msh3 Predict that Mispair Recognition Involves DNA Bending and Strand Separation. *Molecular and Cellular Biology*, *30*(13), 3321–3328. doi: 10.1128/mcb.01558-09
3. Fishel, R. (2015). Mismatch Repair. *The Journal of Biological Chemistry*, *290*(44), 26395–26403. doi: 10.1074/jbc.R115.660142
4. Goellner, E. M., Putnam, C. D., & Kolodner, R. D. (2015). Exonuclease 1-dependent and independent mismatch repair . *Elsevier Science*, *32*, 24–32. doi: 10.1016/j.dnarep.2015.04.010
5. Guo, J., Chen, L., & Li, G.-M. (2017). DNA mismatch repair in trinucleotide repeat instability. *Science China Life Sciences*, *60*(1087), 1087–1092. doi: 10.1007/s11427-017-9186-7
6. Kolodner, R. D., & Marsischky, G. T. (1999). Eukaryotic DNA mismatch repair. *Elsevier Science*, *9*(1), 89–96. doi: 10.1016/S0959-437X(99)80013-6
7. Liu, J., Hanne, J., Britton, B. M., Bennett, J., Kim, D., Lee, J.-B., & Fishel, R. (2016). Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. *Nature*, *539*(7630), 583–587. doi: 10.1038/nature20562
8. Theile, C. S., Witte, M. D., Blom, A. E. M., Kundrat, L., Ploegh, H. L., & Guimaraes, C. P. (2013). Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nature Protocols*, *8*(9), 1800–1807. doi: 10.1038/nprot.2013.102

9. Haraldsdottir, S., Rafnar, T., Frankel, W. L., Einarsdottir, S., Sigurdsson, A., Hampel, H., … Stefansson, K. (2017). Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2. Nature Communications, 8(1). doi: 10.1038/ncomms14755
10. Pringsheim, T., Wilshire, K., Day, L., Dykeman, J., Steeves, T., & Jette, N. (2012). The Incidence and Prevalence of Huntington's Disease: A Systematic Review and Meta-analysis . Movement Disorders, 27(9). doi: 10.1002/mds.25075
11. Schmidt, M. H., & Pearson, C. E. (2016). Disease-associated repeat instability and mismatch repair. DNA Repair, 38, 117–126. doi: 10.1016/j.dnarep.2015.11.008
12. Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D., & Hitchins, M. P. (2015). Milestones of Lynch syndrome: 1895–2015. Nature Reviews Cancer, 15(3), 181–194. doi: 10.1038/nrc3878
13. van den Broek, W. J. A. A., Nelen, M. R., Wansink, D. G., Coerwinkel, M. M., te Riele, H., Groenen, P. J. T. A., & Wieringa, B. (2002). Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. Human Molecular Genetics, 11(2), 191–198. doi: 10.1093/hmg/11.2.191